The Effects of Temperature on the Spring Green-up in a Northern Minnesota Forest: A Statistical Analysis

Thomas Cuezze, Henry Jacobson

Introduction

As the world grows warmer, ecosystems across the world are changing. This effect is especially profound in the boreal parts of the globe, which are widely considered to be particularly vulnerable to climate change. One possible effect of climate change is the earlier onset of spring greenup, the time of year in which vegetation recovers from winter dormancy and "greens up" in preparation for summer. In order to explore the mechanics of these changes, data were collected from 10 experimental enclosures on the Marshall Experimental Forest in northern Minnesota (Richardson et al., 2018). At each enclosure, three different vegetation types were studied: *Picea mariana*, the black spruce, *Larix laricina*, the larch, and a ground level mixed shrub community. All three were present at each of the 10 sites except for one site where no *Larix* was observed, for a total number of observations $n = 29$. Sample size for each group is given in **Table 1.** Each enclosure was randomly assigned to one of five target temperature treatment levels (0°C, +2.25°C, +4.5°C, +6.75°C, or +9°C). Heating elements were used to raise the ambient temperature of the enclosure, and the actual temperature level above the ambient was recorded (in degrees Celsius). The response recorded was the spring greenup day of year: the day of the year (1 to 365) when the vegetation began greenup, in days. Vegetation type and temperature level above ambient were assigned as predictors. The primary research questions are: how does ambient temperature affect date of spring greenup, and how does that effect change across different vegetation types?

Statistical Procedures Used

Figure 1 illustrates the mean spring greenup day of year across groups as well as the distribution of the data and 95% confidence intervals, plotted using an enhanced stripchart from the catstats2 package (Greenwood, 2023) in R (R Core Team, 2023). Looking at the chart, it is obvious that there are differences in mean greenup date across vegetation types. **Table 2**, created using the knitr package, (Xie, 2023) displays the exact sample mean for each group, along with the standard deviation, median, minumum, maximum, and quantiles. *Larix* trees had a mean greenup day of 104.0, mixed shrub had a mean greenup day of 131.0, and *Picea* trees had a mean greenup day of 64.4.

To examine the relationship between temperature and greenup day, we created a scatterplot using the ggplot2 package (Wickham 2016), displayed in **Figure 2.** Linear regression lines for each group have been added to the plot to illustrate the relationship. A strong, negative, and linear relationship between temperature and transition date is observed for all three vegetation type groups. It appears that the slope may vary across groups, however, implying that an interaction term may be needed. To further assess this relationship, we fitted a multiple linear regression (MLR) model of transition date regressed on temperature and transition date using R. A second MLR model with the same predictors as well as an interaction term between them was also fit.

Assumptions for multiple linear regression were assessed using a suite of diagnostic plots created using the ggResidpanel package (Goode and Rey, 2023). In **Figure 3**, we fit the diagnostic plots for the full model (with interaction terms). Looking at the normal Q-Q plot, there is weak evidence against the assumption of normality in the model because the residuals in the normal Q-Q plot vary only slightly from the one-to-one line. Next, in the residuals vs. fitted plot, there is weak evidence against the assumption of constant variance because the vertical spread of the residuals does not change substantially across the fitted values. There is also weak evidence against the assumption of linearity because the residuals in the R v. F plot do not display any noticeable curvature. Finally, looking at the Residual vs. Leverage plot, there do not appear to be any problems with influential observations as none of the points have a Cook's Distance greater than 0.5.

There is a potential problem with the independence of observations, as each of the enclosures was located in one of three blocks, and it is not possible to know that conditions were identical across the three blocks. Additionally, the observations are taken from a limited number of enclosures, and each enclosure shares three observations (one of each vegetation type), which could cause observation outcomes to be non-independent – for example, if the plants are shading each other, signalling to each other, or sharing/stealing resources. For this reason, we should be cautious about any conclusions drawn from the model. Assessing the additive model, all of the residual plots (given in **Figure 4**) looked similar and showed similar evidence for each of the assumptions; with the exception of the possible problems with independence of observations, we conclude that multiple linear regression is an appropriate modeling tool for these data and no transformations are necessary.

Summary of Statistical Findings

The final estimated model was: $\hat{\mu}\{\text{Transition Day of Year}\} = 112.57 - 1.32(\text{Temperature}) + I_{\text{Vegetation Type = Mixed Shrub}}(26.11) + I_{\text{Vegetation Type = Picea}}(-40.49)$. $I_{\text{Vegetation Type = Mixed Shrub}}$ is an indicator variable equal to 1 when the vegetation type is mixed shrub and 0 in all other cases, and $I_{\text{Vegetation Type = Picea}}$ is an indicator variable equal to 1 when the vegetation type is mixed shrub and 0 in all other cases. The model uses *Larix* as the reference level. This is the additive model.

The interaction model was also considered, and compared to the additive model with Type II ANOVA, using the car package in R (Fox et al., 2022). A Type II sums-of-squares F-test provided weak evidence for the null hypothesis of no interaction between temperature and vegetation type on spring greenup date ($F_{2,23} = 1.249$, p-value = .3055), so we conclude that there is no interaction between temperature and vegetation type and we should proceed with the additive model. Type II sums-of-squares F-tests were also conducted for all of the terms in the additive model. An ANOVA table for all predictors in the model can be found in **Table 3.** There is strong evidence against the null hypothesis of no effect of temperature on spring greenup date ($F_{1,23} = 104.43$, p-value < .001), so we conclude that temperature has an effect on spring greenup date. There is also strong evidence against the null hypothesis of no effect of vegetation type on spring greenup date ($F_{2,23} = 1567.405$, p-value < .001), so we conclude that vegetation type has an effect on spring greenup date.

It is estimated that for a 1 degree Celsius increase in temperature, the true mean greenup day of year decreases by 1.6834 days after controlling for vegetation type (95% CI: -2.026059 to -1.340787). In **Figure 5**, effects plots were created using the effects package (Fox et al., 2022) to visualize this relationship. This relationship did not change across vegetation types.

Scope of Inference

In the experimental design, the enclosure sites were not randomly sampled from any particular population. Therefore, it is not possible to generalize the results of this study beyond the 10 Marshall Experimental Forest sites included in the study. The study still provides useful information about this site and sites judged to be similar, but we cannot assume the same statistical relationships will hold for other sites. However, temperature was randomly assigned for each enclosure across the 10 study sites, therefore we can infer a causative effect of temperature on spring greenup date.

We can conclude that increased temperatures above the ambient cause an earlier spring greenup date, and that this relationship does not vary across vegetation types. We can also conclude that vegetation type has a causal effect on spring greenup date. That said, in light of the specified scope of inference, these conclusions are confined to the 10 Marshall Experimental Forest sites studied from Autumn 2015 to Spring 2016. Further exploration could delve into the nuances of this relationship with a randomized sample of enclosure sites.

Appendix

References

1. Richardson, A.D., Hufkens, K., Milliman, T. et al. Ecosystem warming extends vegetation activity but heightens vulnerability to cold temperatures. Nature 560, 368–371 (2018). https://doi.org/10.1038/s41586-018-0399-1
2. Greenwood, M (2023). *catstats2: Upper Level Statistics for Montana State University Bobcats*. R package version 0.1.
3. R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
4. Xie, Y (2023). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.43, https://yihui.org/knitr/.
5. Wickham, H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.
6. Goode, K, Rey, K (2023). *ggResidpanel: Panels and Interactive Versions of Diagnostic Plots using 'ggplot2'*. R package version 0.3.0.9000, https://goodekat.github.io/ggResidpanel/.
7. Fox, J, Weisberg, S, Price, B (2022). *carData: Companion to Applied Regression Data Sets*. R package version 3.0-5, https://CRAN.R-project.org/package=carData.

Figure 1: Enhanced Stripchart

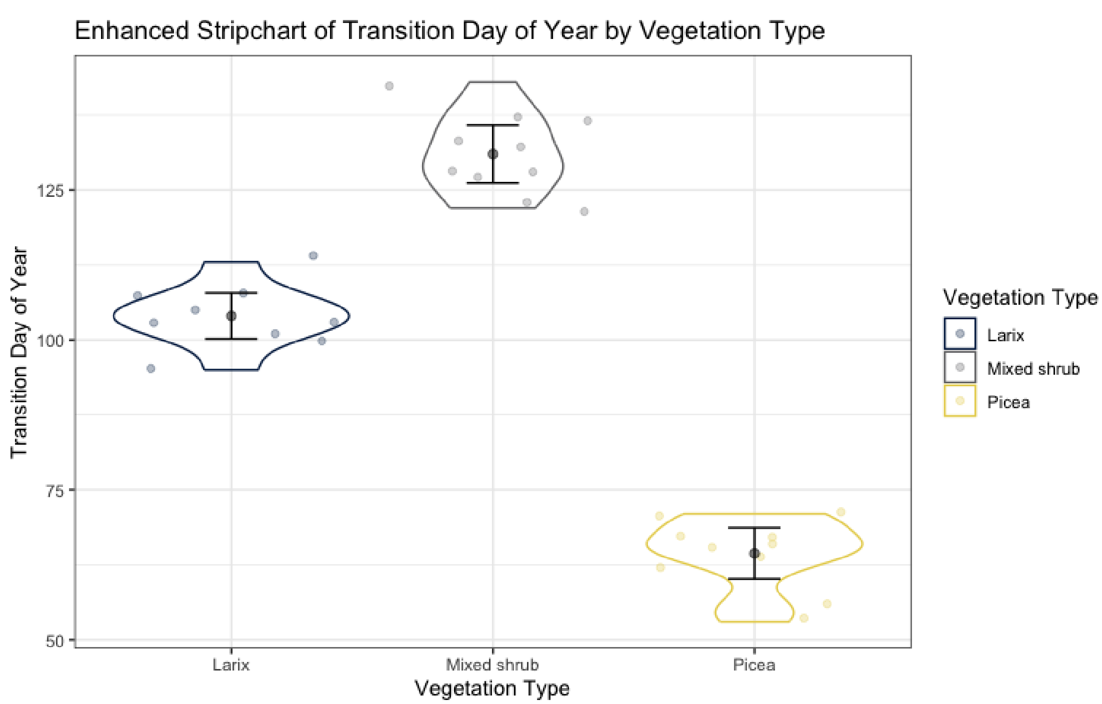


Figure 2: Scatterplot

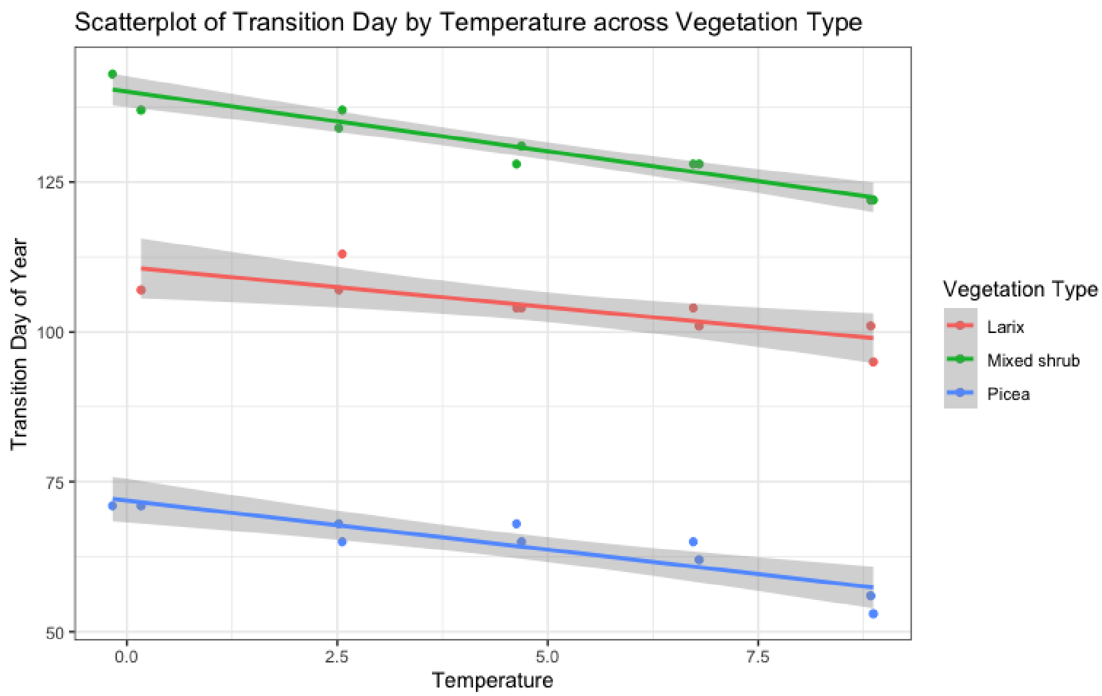


Figure 3: Diagnostic Panel for Full Model

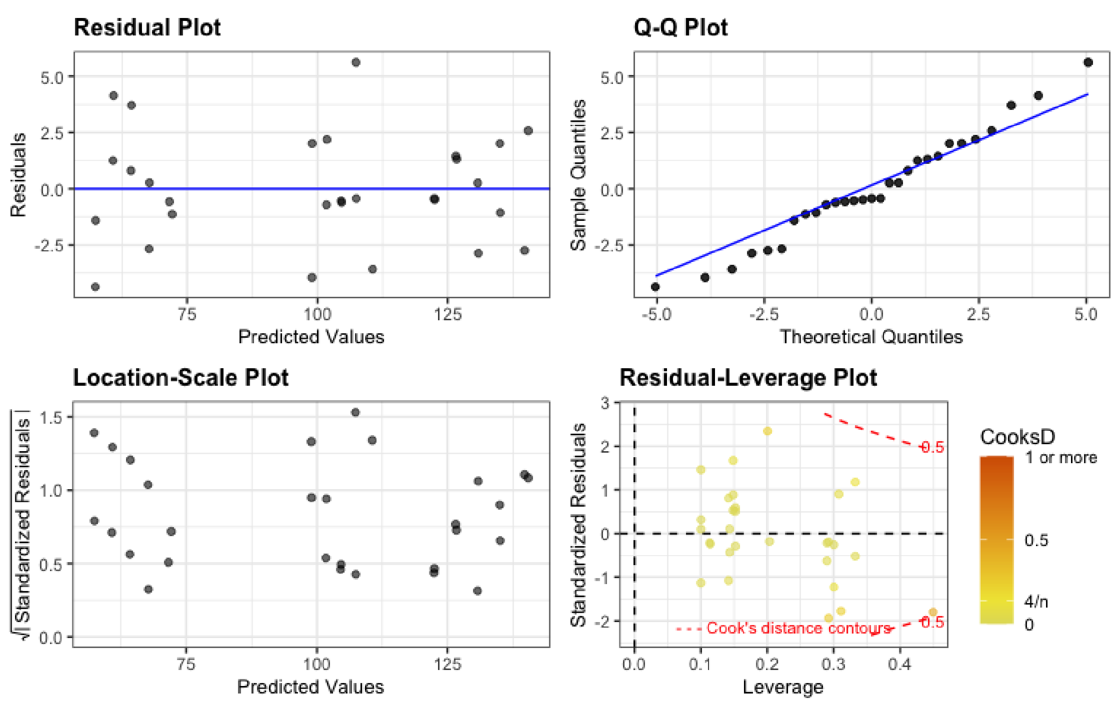


Figure 4: Diagnostic Panel for Additive Model

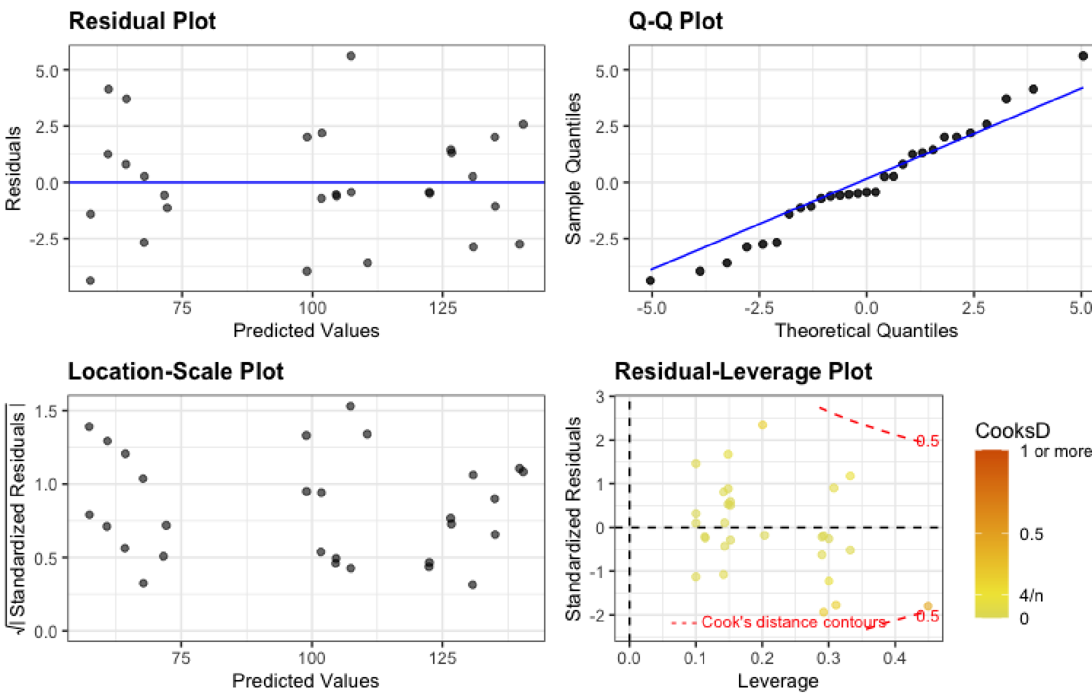


Figure 5: Effects Plot

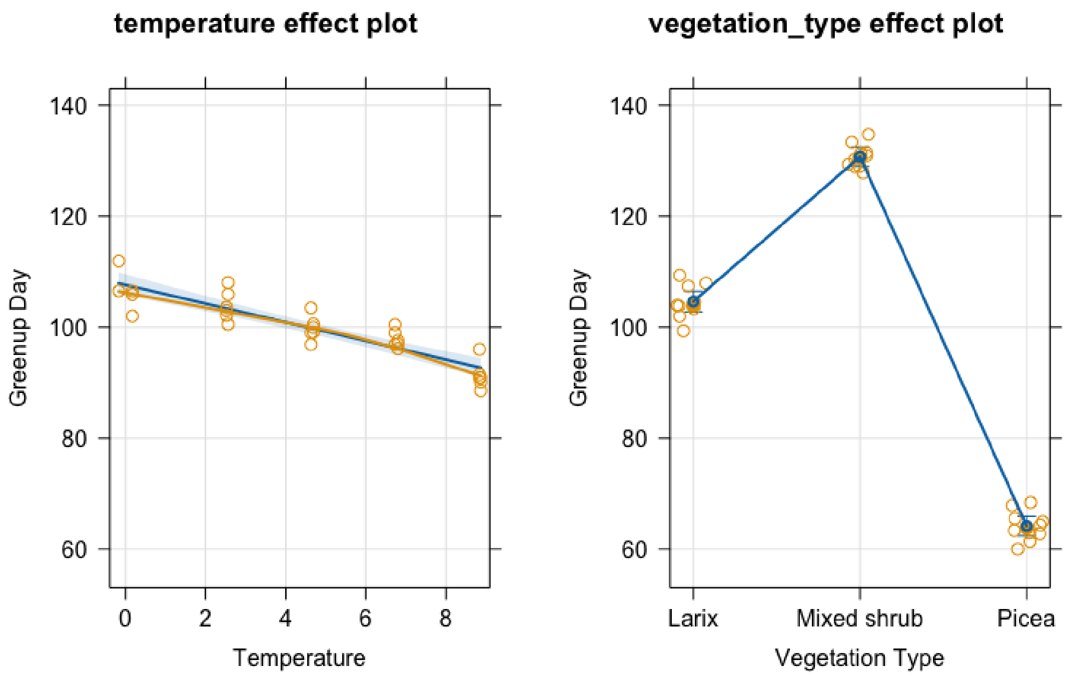


Table 1: Sample Sizes for Subgroups

Vegetation Type	Sample Size (n)
Larix	9
Mixed Shrub	10
Picea	10
Total	29

Table 2: Summary Statistics

Vegetation Type	Mean Transition Day	Minimum Transition Day	25th Percentile Transition Day	Median Transition Day	75th Percentile Transition Day	Maximum Transition Day	Transition Day Standard Deviation
Larix	104.0	95	101.00	104.0	107.00	113	4.974937
Mixed shrub	131.0	122	128.00	129.5	136.25	143	6.782330
Picea	64.4	53	62.75	65.0	68.00	71	5.966574

Table 3: Type II ANOVA for the Additive Model

Predictor	Sums of Squares	Degrees of Freedom	F Statistic	P-Value
Temperature	749.4196	1	102.3907	<0.001
Vegetation Type	22496.2014	2	1536.7902	<0.001
Residuals	182.9804	25	NA	NA

The Effects of Temperature on the Spring Green-up in a Northern Minnesota Forest: A Statistical Analysis

Thomas Cuezze, Henry Jacobson

Introduction

As the world grows warmer, ecosystems across the world are changing. This effect is especially profound in the boreal parts of the globe, which are widely considered to be particularly vulnerable to climate change. One possible effect of climate change is the earlier onset of spring greenup, the time of year in which vegetation recovers from winter dormancy and "greens up" in preparation for summer. In order to explore the mechanics of these changes, data were collected from 10 experimental enclosures on the Marshall Experimental Forest in northern Minnesota (Richardson et al., 2018). At each enclosure, three different vegetation types were studied: *Picea mariana*, the black spruce, *Larix laricina*, the larch, and a ground level mixed shrub community. All three were present at each of the 10 sites except for one site where no *Larix* was observed, for a total number of observations $n = 29$. Sample size for each group is given in **Table 1.** Each enclosure was randomly assigned to one of five target temperature treatment levels (0°C, +2.25°C, +4.5°C, +6.75°C, or +9°C). Heating elements were used to raise the ambient temperature of the enclosure, and the actual temperature level above the ambient was recorded (in degrees Celsius). The response recorded was the spring greenup day of year: the day of the year (1 to 365) when the vegetation began greenup, in days. Vegetation type and temperature level above ambient were assigned as predictors. The primary research questions are: how does ambient temperature affect date of spring greenup, and how does that effect change across different vegetation types?

Statistical Procedures Used

Figure 1 illustrates the mean spring greenup day of year across groups as well as the distribution of the data and 95% confidence intervals, plotted using an enhanced stripchart from the catstats2 package (Greenwood, 2023) in R (R Core Team, 2023). Looking at the chart, it is obvious that there are differences in mean greenup date across vegetation types. **Table 2**, created using the knitr package, (Xie, 2023) displays the exact sample mean for each group, along with the standard deviation, median, minumum, maximum, and quantiles. *Larix* trees had a mean greenup day of 104.0, mixed shrub had a mean greenup day of 131.0, and *Picea* trees had a mean greenup day of 64.4.

To examine the relationship between temperature and greenup day, we created a scatterplot using the ggplot2 package (Wickham 2016), displayed in **Figure 2.** Linear regression lines for each group have been added to the plot to illustrate the relationship. A strong, negative, and linear relationship between temperature and transition date is observed for all three vegetation type groups. It appears that the slope may vary across groups, however, implying that an interaction term may be needed. To further assess this relationship, we fitted a multiple linear regression (MLR) model of transition date regressed on temperature and transition date using R. A second MLR model with the same predictors as well as an interaction term between them was also fit.

Assumptions for multiple linear regression were assessed using a suite of diagnostic plots created using the ggResidpanel package (Goode and Rey, 2023). In **Figure 3**, we fit the diagnostic plots for the full model (with interaction terms). Looking at the normal Q-Q plot, there is weak evidence against the assumption of normality in the model because the residuals in the normal Q-Q plot vary only slightly from the one-to-one line. Next, in the residuals vs. fitted plot, there is weak evidence against the assumption of constant variance because the vertical spread of the residuals does not change substantially across the fitted values. There is also weak evidence against the assumption of linearity because the residuals in the R v. F plot do not display any noticeable curvature. Finally, looking at the Residual vs. Leverage plot, there do not appear to be any problems with influential observations as none of the points have a Cook's Distance greater than 0.5.

There is a potential problem with the independence of observations, as each of the enclosures was located in one of three blocks, and it is not possible to know that conditions were identical across the three blocks. Additionally, the observations are taken from a limited number of enclosures, and each enclosure shares three observations (one of each vegetation type), which could cause observation outcomes to be non-independent – for example, if the plants are shading each other, signalling to each other, or sharing/stealing resources. For this reason, we should be cautious about any conclusions drawn from the model. Assessing the additive model, all of the residual plots (given in **Figure 4**) looked similar and showed similar evidence for each of the assumptions; with the exception of the possible problems with independence of observations, we conclude that multiple linear regression is an appropriate modeling tool for these data and no transformations are necessary.

Summary of Statistical Findings

The final estimated model was: $\hat{\mu}\{\text{Transition Day of Year}\} = 112.57 - 1.32(\text{Temperature}) + I_{\text{Vegetation Type = Mixed Shrub}}(26.11) + I_{\text{Vegetation Type = Picea}}(-40.49)$. $I_{\text{Vegetation Type = Mixed Shrub}}$ is an indicator variable equal to 1 when the vegetation type is mixed shrub and 0 in all other cases, and $I_{\text{Vegetation Type = Picea}}$ is an indicator variable equal to 1 when the vegetation type is mixed shrub and 0 in all other cases. The model uses *Larix* as the reference level. This is the additive model.

The interaction model was also considered, and compared to the additive model with Type II ANOVA, using the car package in R (Fox et al., 2022). A Type II sums-of-squares F-test provided weak evidence for the null hypothesis of no interaction between temperature and vegetation type on spring greenup date ($F_{2,23} = 1.249$, p-value = .3055), so we conclude that there is no interaction between temperature and vegetation type and we should proceed with the additive model. Type II sums-of-squares F-tests were also conducted for all of the terms in the additive model. An ANOVA table for all predictors in the model can be found in **Table 3.** There is strong evidence against the null hypothesis of no effect of temperature on spring greenup date ($F_{1,23} = 104.43$, p-value < .001), so we conclude that temperature has an effect on spring greenup date. There is also strong evidence against the null hypothesis of no effect of vegetation type on spring greenup date ($F_{2,23} = 1567.405$, p-value < .001), so we conclude that vegetation type has an effect on spring greenup date.

It is estimated that for a 1 degree Celsius increase in temperature, the true mean greenup day of year decreases by 1.6834 days after controlling for vegetation type (95% CI: -2.026059 to -1.340787). In **Figure 5**, effects plots were created using the effects package (Fox et al., 2022) to visualize this relationship. This relationship did not change across vegetation types.

Scope of Inference

In the experimental design, the enclosure sites were not randomly sampled from any particular population. Therefore, it is not possible to generalize the results of this study beyond the 10 Marshall Experimental Forest sites included in the study. The study still provides useful information about this site and sites judged to be similar, but we cannot assume the same statistical relationships will hold for other sites. However, temperature was randomly assigned for each enclosure across the 10 study sites, therefore we can infer a causative effect of temperature on spring greenup date.

We can conclude that increased temperatures above the ambient cause an earlier spring greenup date, and that this relationship does not vary across vegetation types. We can also conclude that vegetation type has a causal effect on spring greenup date. That said, in light of the specified scope of inference, these conclusions are confined to the 10 Marshall Experimental Forest sites studied from Autumn 2015 to Spring 2016. Further exploration could delve into the nuances of this relationship with a randomized sample of enclosure sites.

Appendix

References

1. Richardson, A.D., Hufkens, K., Milliman, T. et al. Ecosystem warming extends vegetation activity but heightens vulnerability to cold temperatures. Nature 560, 368–371 (2018). https://doi.org/10.1038/s41586-018-0399-1
2. Greenwood, M (2023). *catstats2: Upper Level Statistics for Montana State University Bobcats*. R package version 0.1.
3. R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

4. Xie, Y (2023). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.43, https://yihui.org/knitr/.
5. Wickham, H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.
6. Goode, K, Rey, K (2023). *ggResidpanel: Panels and Interactive Versions of Diagnostic Plots using 'ggplot2'*. R package version 0.3.0.9000, https://goodekat.github.io/ggResidpanel/.
7. Fox, J, Weisberg, S, Price, B (2022). *carData: Companion to Applied Regression Data Sets*. R package version 3.0-5, https://CRAN.R-project.org/package=carData.


```
##R Output 1
lm_greenup <- lm(transition_date ~ temperature * vegetation_type, data =
greenup_2016)
summary(lm_greenup)

##
## Call:
## lm(formula = transition_date ~ temperature * vegetation_type,
##     data = greenup_2016)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3654 -1.1338 -0.4354  1.4505  5.6181
##
## Coefficients:
##                                        Estimate Std. Error t value
Pr(>|t|)
## (Intercept)                            110.8039     1.8428  60.128  < 2e-
16
## temperature                             -1.3367     0.3167  -4.221
0.000325
## vegetation_typeMixed shrub              29.2802     2.3798  12.304 1.34e-
11
## vegetation_typePicea                   -38.9478     2.3798 -16.366 3.63e-
14
## temperature:vegetation_typeMixed shrub  -0.6537     0.4180  -1.564
0.131506
## temperature:vegetation_typePicea        -0.2970     0.4180  -0.710
0.484569
##
## Residual standard error: 2.679 on 23 degrees of freedom
## Multiple R-squared:  0.9929, Adjusted R-squared:  0.9914
## F-statistic: 646.3 on 5 and 23 DF,  p-value: < 2.2e-16

lm_greenup_reduced <- lm(transition_date ~ temperature + vegetation_type,
data = greenup_2016)
summary(lm_greenup_reduced)

##
## Call:
```

```
## lm(formula = transition_date ~ temperature + vegetation_type,
##     data = greenup_2016)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.2824 -1.3970 -0.4409  1.3641  4.7409
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                112.5686     1.2371   91.00  < 2e-16
## temperature                 -1.6834     0.1664  -10.12 2.52e-10
## vegetation_typeMixed shrub  26.1145     1.2461   20.96  < 2e-16
## vegetation_typePicea       -40.4855     1.2461  -32.49  < 2e-16
##
## Residual standard error: 2.705 on 25 degrees of freedom
## Multiple R-squared:  0.9922, Adjusted R-squared:  0.9912
## F-statistic:  1055 on 3 and 25 DF,  p-value: < 2.2e-16
```

```r
##R Output 2
anova(lm_greenup, lm_greenup_reduced)
```

```
## Analysis of Variance Table
##
## Model 1: transition_date ~ temperature * vegetation_type
## Model 2: transition_date ~ temperature + vegetation_type
##   Res.Df    RSS Df Sum of Sq     F Pr(>F)
## 1     23 165.05
## 2     25 182.98 -2   -17.927 1.249 0.3055
```

Figure 1: Enhanced Stripchart

```r
##Figure 1
enhanced_stripchart(data = greenup_2016, transition_date ~ vegetation_type) +
labs(x = "Vegetation Type", y = "Transition Day of Year", title = "Enhanced
Stripchart of Transition Day of Year by Vegetation Type", colour =
"Vegetation Type")
```

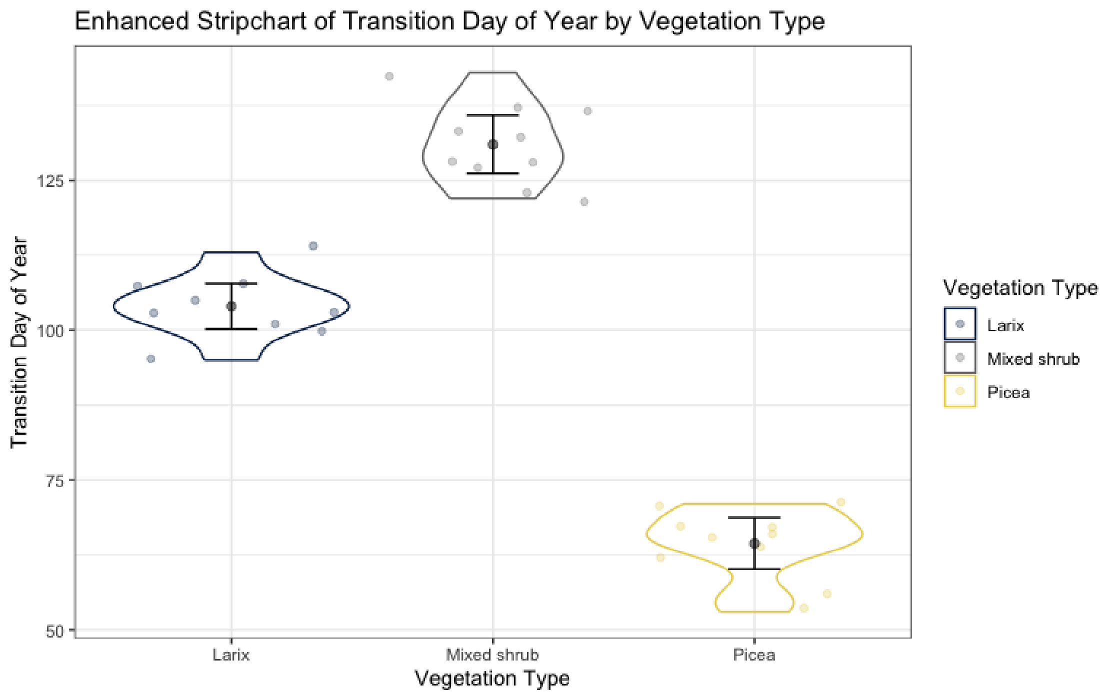


Figure 2: Scatterplot

```
##Figure 2
greenup_2016 %>%
  ggplot(aes(x = temperature, y = transition_date, colour = vegetation_type))
+
  geom_point() +
  labs(x = "Temperature", y = "Transition Day of Year", colour = "Vegetation
Type", title = "Scatterplot of Transition Day by Temperature across
Vegetation Type") +
  geom_smooth(method = "lm")
```

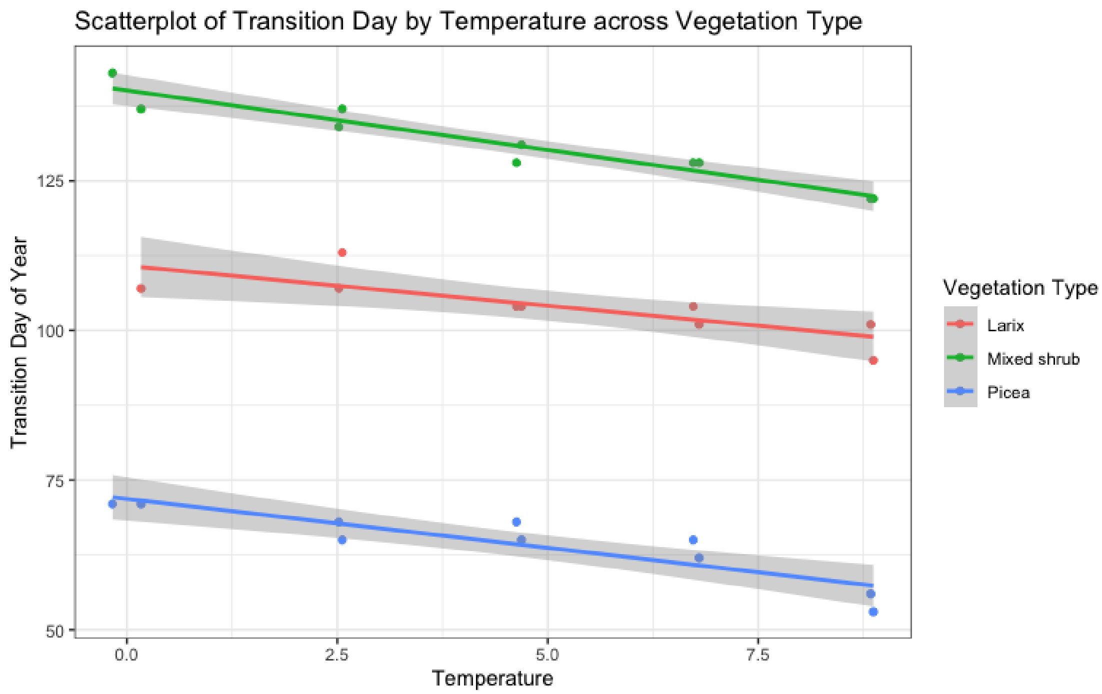


Figure 3: Diagnostic Panel for Full Model

```
##Figure 3
resid_panel(lm_greenup, "R")
```

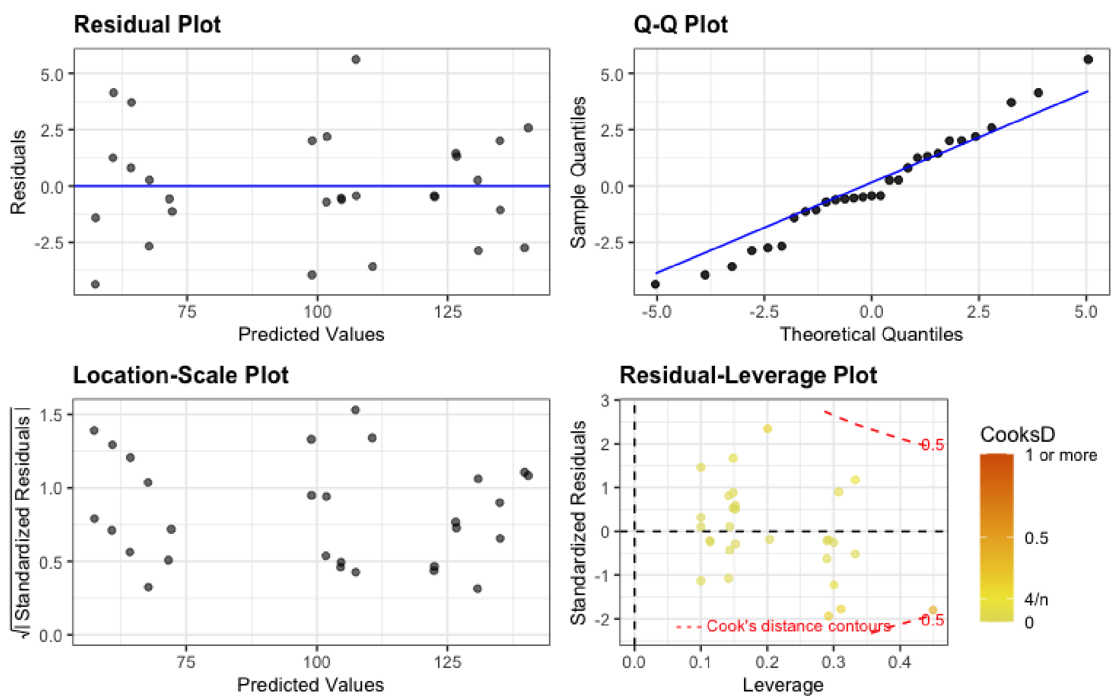


Figure 4: Diagnostic Panel for Additive Model

```
##Figure 4
resid_panel(lm_greenup, "R")
```

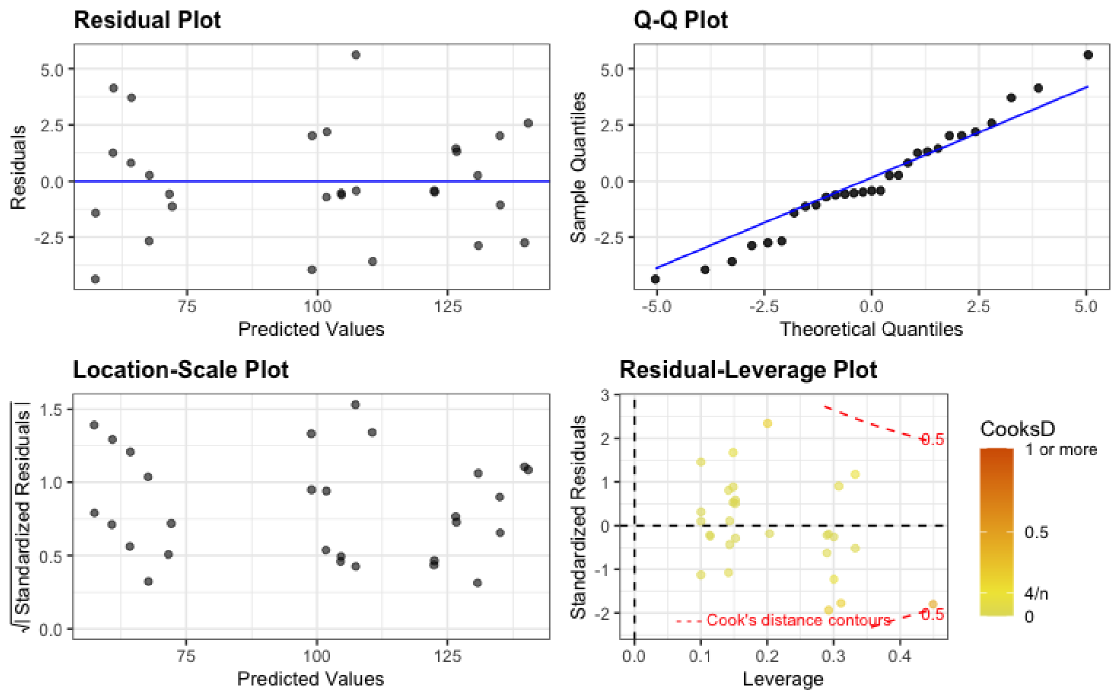
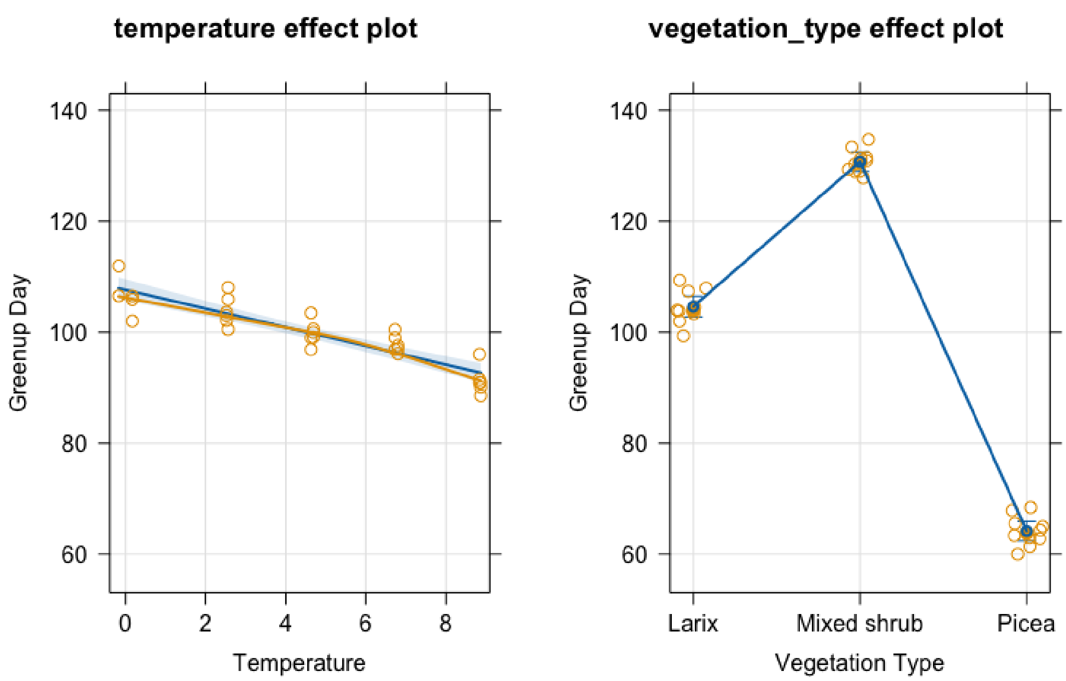


Figure 5: Effects Plot

```
plot(allEffects(lm_greenup_reduced, residuals = T), multiline = T, grid = T,
confint = list(style = "auto"), lty = 1:2, ylab = "Greenup Day", xlab =
list(temperature ="Temperature", vegetation_type ="Vegetation Type"))
```



##Table 1: Sample Sizes for Subgroups

```r
##Table 1
sample_size <- data.frame(c("Larix", "Mixed Shrub", "Picea", "Total"), c(9,
10, 10, 29))
kable(sample_size, col.names = c("Vegetation Type", "Sample Size (n)"))
```

Vegetation Type	Sample Size (n)
Larix	9
Mixed Shrub	10
Picea	10
Total	29

Table 2: Summary Statistics

```r
##Define functions to get first and third quantiles
fq <- function(data) {
  return(quantile(data, probs = .25))
}
tq <- function(data) {
  return(quantile(data, probs = .75))
}
##Table 2
greenup_2016 %>%
  group_by(vegetation_type) %>%
  select(vegetation_type, transition_date) %>%
  summarise_at(vars(transition_date), list(mean = mean, min = min, fq = fq,
median = median, tq = tq, max = max, sd = sd)) %>%
  kable(col.names = c("Vegetation Type", "Mean Transition Day","Minimum
Transition Day", "25th Percentile Transition Day", "Median Transition
Day","75th Percentile Transition Day", "Maximum Transition Day", "Transition
Day Standard Deviation"))
```

Vegetation Type	Mean Transition Day	Minimum Transition Day	25th Percentile Transition Day	Median Transition Day	75th Percentile Transition Day	Maximum Transition Day	Transition Day Standard Deviation
Larix	104.0	95	101.00	104.0	107.00	113	4.974937
Mixed shrub	131.0	122	128.00	129.5	136.25	143	6.782330
Picea	64.4	53	62.75	65.0	68.00	71	5.966574

Table 3: Type II ANOVA for the Additive Model

```r
##Table 2
kable(Anova(lm_greenup_reduced), col.names = c("Sums of Squares", "Degrees of
Freedom", "F Statistic", "P-Value"))
```

	Sums of Squares	Degrees of Freedom	F Statistic	P-Value
temperature	749.4196	1	102.3907	0
vegetation_type	22496.2014	2	1536.7902	0
Residuals	182.9804	25	NA	NA